

# Introduction to pandas for marketing

ANALYZING MARKETING CAMPAIGNS WITH PANDAS



**Jill Rosok**  
Data Scientist

# What does a data scientist on a marketing team do?



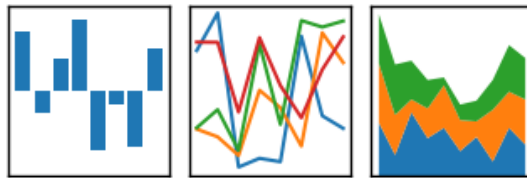
- Analyzing marketing campaign performance
- Attributing credit for conversions to marketing channels
- A/B testing

# What is pandas, again?

- Provides table-like data structures that are easy to use in analysis
- Allows for easy importing and exporting of a variety of common formats (i.e., CSV, TSV, Stata)
- Enables manipulation such as joining other datasets, grouping by and aggregating columns, and taking subsets of dataset columns and rows.

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



# Importing data using pandas

```
import pandas as pd
```

```
marketing = pd.read_csv('marketing.csv')
```

# Inspecting data

```
print(marketing.head())
```

```
   user_id date_served channel variant conv \
0  a100000029  2018-01-01 House Ads personalization True
1  a100000030  2018-01-01 House Ads personalization True
2  a100000031  2018-01-01 House Ads personalization True
3  a100000032  2018-01-01 House Ads personalization True
4  a100000033  2018-01-01 House Ads personalization True

   language_displayed language_preferred age_group
0           English           English  0-18 years
1           English           English  19-24 years
2           English           English  24-30 years
3           English           English  30-36 years
4           English           English  36-45 years
```

# Summary statistics

```
print(marketing.describe())
```

```
      user_id  date_served  channel  variant  conv  \
count      9882      9881      9882      9882  9882
unique      7253        31         5         2     2
top  a100000882  2018-01-15  House Ads  control  False
freq         6        782      4682      4986  8883

      language_displayed  language_preferred  age_group
count              9882              9882      9882
unique               4                4         7
top              English              English  19-24 years
freq             9695              9177      1650
```

# Missing values & data types

```
print(marketing.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9996 entries, 0 to 9995
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
#   -----
0   user_id                9996 non-null   object
1   date_served            9980 non-null   object
...
9   date_subscribed        1815 non-null   object
10  date_canceled           568 non-null    object
11  subscribing_channel     1815 non-null   object
12  is_retained             1815 non-null   object
dtypes: object(12)
memory usage: 937.2+ KB
```

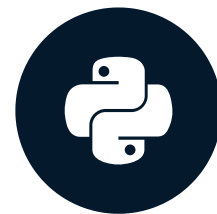
# Let's Practice!

ANALYZING MARKETING CAMPAIGNS WITH PANDAS



# Data types and data merging

ANALYZING MARKETING CAMPAIGNS WITH PANDAS



**Jill Rosok**  
Data Scientist

# Common data types

- Strings (objects)
- Numbers (floats, integers)
- Boolean values (True, False)
- Dates

# Data type of a column

```
# Print a data type of a single column  
print(marketing['converted'].dtype)
```

```
dtype('object')
```

# Changing the data type of a column

```
# Change the data type of a column
marketing['converted'] = marketing['converted']\
                        .astype('bool')
print(marketing['converted'].dtype)
```

```
dtype('bool')
```

# Creating new boolean columns

```
marketing['is_house_ads'] = np.where(
    marketing['marketing_channel'] == 'House Ads',
    True, False
)

print(marketing.is_house_ads.head(3))
```

```
0    True
1   False
2    True
Name: is_house_ads, dtype: bool
```

# Mapping values to existing columns

```
channel_dict = {"House Ads": 1, "Instagram": 2,  
               "Facebook": 3, "Email": 4, "Push": 5}  
marketing['channel_code'] = marketing['marketing_channel']\  
    .map(channel_dict)  
  
print(marketing['channel_code'].head(3))
```

```
0    1  
1    1  
2    1  
Name: channel_code, dtype: int64
```

# Date columns

```
# Read date columns using parse_dates
marketing = pd.read_csv('marketing.csv',
                        parse_dates=['date_served',
                                    'date_subscribed',
                                    'date_canceled'])

# Or
# Convert already existing column to datetime column
marketing['date_served'] = pd.to_datetime(
    marketing['date_served']
)
```

# Date columns

```
# Or convert each column individually
# Convert already existing column to datetime column
marketing['date_served'] = pd.to_datetime(
    marketing['date_served']
)
```



# Date columns

```
marketing['day_served'] = marketing['date_served']\
    .dt.dayofweek
```

# Let's Practice!

ANALYZING MARKETING CAMPAIGNS WITH PANDAS

# Initial exploratory analysis

ANALYZING MARKETING CAMPAIGNS WITH PANDAS



**Jill Rosok**  
Data Scientist

# How many users see marketing assets?

```
# Aggregate unique users that see ads by date
daily_users = marketing.groupby(['date_served'])\
                        ['user_id'].nunique()

print(daily_users)
```

```
date_served
2018-01-01    362
2018-01-02    374
2018-01-03    348
...
Name: user_id, dtype: int64
```

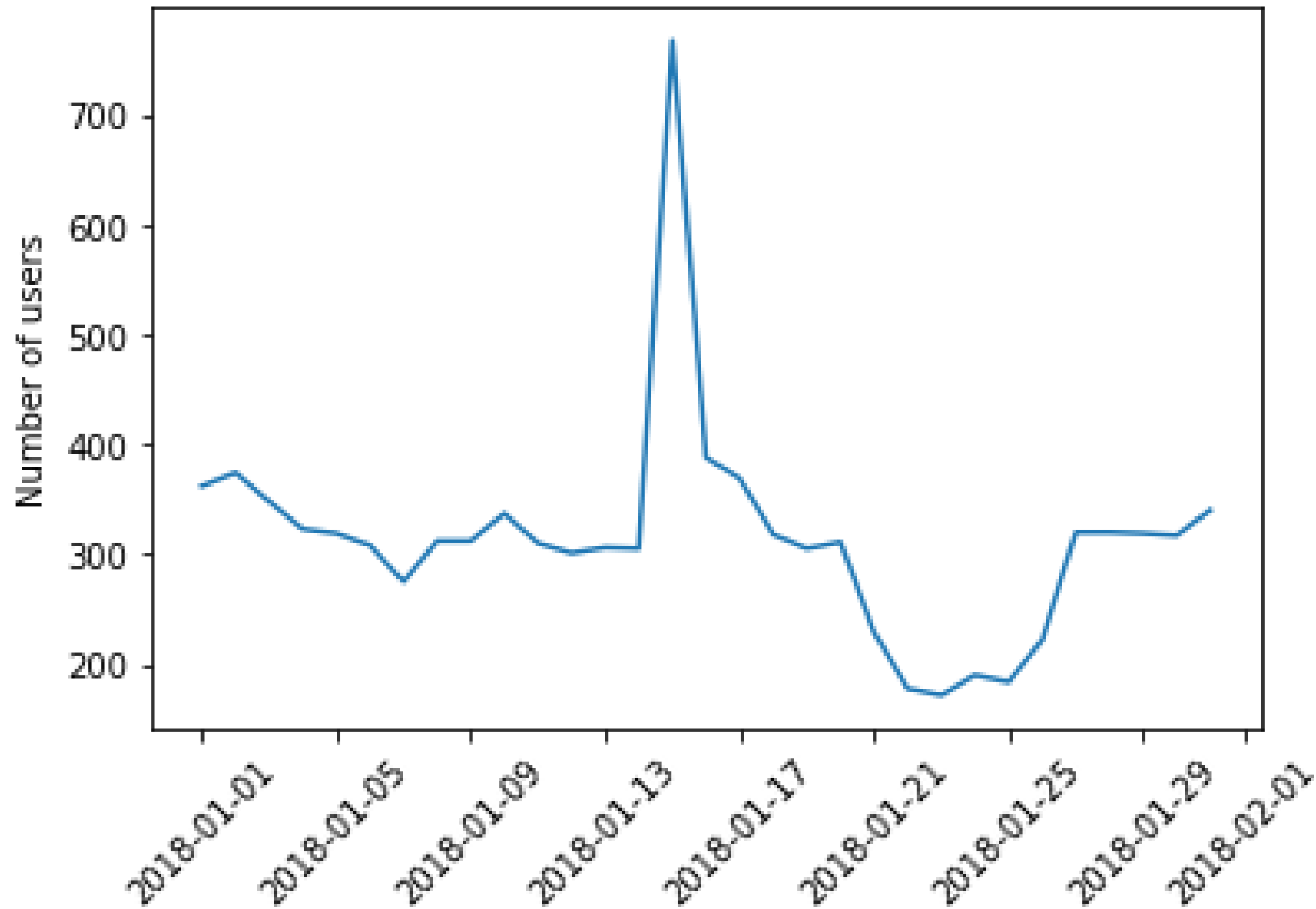
# Visualizing results

```
import matplotlib.pyplot as plt

# Plot
daily_users.plot()

# Annotate
plt.title('Daily number of users who see ads')
plt.xlabel('Date')
plt.ylabel('Number of users')
plt.xticks(rotation = 45)
plt.show()
```

Daily number of users who see ads



# Let's practice!

ANALYZING MARKETING CAMPAIGNS WITH PANDAS